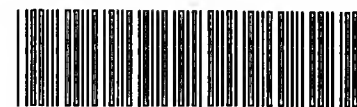


0570  
0807

# 12



ENTERED

OIFE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/995,860

DATE: 08/09/2002

TIME: 14:42:30

Input Set : A:\SEQUENCE LSITING.txt

Output Set: N:\CRF4\08092002\I995860.raw

and

3 <110> APPLICANT: Innogenetics N.V.

5 <120> TITLE OF INVENTION: Purified hepatitis C virus envelope proteins for diagnostic

6 therapeutic use.

8 <130> FILE REFERENCE: 2551-69

10 <140> CURRENT APPLICATION NUMBER: 09/995,860

C--> 12 <141> CURRENT FILING DATE: 2001-11-29

12 <160> NUMBER OF SEQ ID NOS: 122

14 <170> SOFTWARE: PatentIn 3.1

16 <210> SEQ ID NO: 1

18 <211> LENGTH: 21

19 <212> TYPE: DNA

20 <213> ORGANISM: Hepatitis C virus

22 <400> SEQUENCE: 1

24 ggcatgcaag cttaattaat t 21

27 <210> SEQ ID NO: 2

29 <211> LENGTH: 68

30 <212> TYPE: DNA

31 <213> ORGANISM: Hepatitis C virus

33 <400> SEQUENCE: 2

35 ccggggaggc ctgcacgtga tcgagggcag acaccatcac caccatcact aatagttaat 60

37 taactgca 68

40 <210> SEQ ID NO: 3

42 <211> LENGTH: 642

43 <212> TYPE: DNA

44 <213> ORGANISM: Hepatitis C virus

46 <220> FEATURE:

47 <221> NAME/KEY: CDS

48 <222> LOCATION: 1..639

50 <220> FEATURE:

51 <221> NAME/KEY: mat\_peptide

52 <222> LOCATION: 1..636

55 <400> SEQUENCE: 3

57 atg ccc ggt tgc tct ttc tct atc ttc ctc ttg gct tta ctg tcc tgt 48

58 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Ala Leu Leu Ser Cys

59 1 5 10 15

61 ctg acc att cca gct tcc gct tat gag gtg cgc aac gtg tcc ggg atg 96

62 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met

63 20 25 30

65 tac cat gtc acg aac gac tgc tcc aac tca agc att gtg tat gag gca 144

66 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala

67 35 40 45

69 gcg gac atg atc atg cac acc ccc ggg tgc gtg ccc tgc gtt cgg gag 192

70 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/995,860

DATE: 08/09/2002

TIME: 14:42:30

Input Set : A:\SEQUENCE LISTING.txt

Output Set: N:\CRF4\08092002\I995860.raw

```

71      50      55      60
73 aac aac tct tcc cgc tgc tgg gta gcg ctc acc ccc acg ctc gca gct      240
74 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
75 65      70      75      80
77 agg aac gcc agc gtc ccc acc acg aca ata cga cgc cac gtc gat ttg      288
78 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
79      85      90      95
81 ctc gtt ggg gcg gct gct ctc tgt tcc gct atg tac gtg ggg gat ctc      336
82 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
83      100      105      110
85 tgc gga tct gtc ttc ctc gtc tcc cag ctg ttc acc atc tcg cct cgc      384
86 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
87      115      120      125
89 cgg cat gag acg gtg cag gac tgc aat tgc tca atc tat ccc ggc cac      432
90 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
91      130      135      140
93 ata aca ggt cac cgt atg gct tgg gat atg atg atg aac tgg tcg cct      480
94 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
95 145      150      155      160
97 aca acg gcc ctg gtg gta tcg cag ctg ctc cgg atc cca caa gct gtc      528
98 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
99      165      170      175
101 gtg gac atg gtg gcg ggg gcc cat tgg gga gtc ctg gcg ggc ctc gcc      576
102 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
103      180      185      190
105 tac tat tcc atg gtg ggg aac tgg gct aag gtt ttg att gtg atg cta      624
106 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
107      195      200      205
109 ctc ttt gct ctc taatag      642
110 Leu Phe Ala Leu
111      210
114 <210> SEQ ID NO: 4
116 <211> LENGTH: 212
117 <212> TYPE: PRT
118 <213> ORGANISM: Hepatitis C virus
121 <400> SEQUENCE: 4
123 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
124 1      5      10      15
126 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
127      20      25      30
129 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
130      35      40      45
132 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
133      50      55      60
135 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
136 65      70      75      80
138 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
139      85      90      95
141 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu

```

## RAW SEQUENCE LISTING

DATE: 08/09/2002

PATENT APPLICATION: US/09/995,860

TIME: 14:42:30

Input Set : A:\SEQUENCE LISTING.txt

Output Set: N:\CRF4\08092002\I995860.raw

```

142          100          105          110
144 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
145          115          120          125
147 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
148          130          135          140
150 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
151 145          150          155          160
153 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
154          165          170          175
156 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
157          180          185          190
159 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
160          195          200          205
162 Leu Phe Ala Leu
163          210
166 <210> SEQ ID NO: 5
168 <211> LENGTH: 795
169 <212> TYPE: DNA
170 <213> ORGANISM: Hepatitis C virus
172 <220> FEATURE:
173 <221> NAME/KEY: CDS
174 <222> LOCATION: 1..792
176 <220> FEATURE:
177 <221> NAME/KEY: mat_peptide
178 <222> LOCATION: 1..789
180 <400> SEQUENCE: 5
182 atg ttg ggt aag gtc atc gat acc ctt aca tgc ggc ttc gcc gac ctc      48
183 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
184 1          5          10          15
186 gtg ggg tac att ccg ctc gtc ggc gcc ccc cta ggg ggc gct gcc agg      96
187 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
188          20          25          30
190 gcc ctg gcg cat ggc gtc cgg gtt ctg gag gac ggc gtg aac tat gca      144
191 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
192          35          40          45
194 aca ggg aat ttg ccc ggt tgc tct ttc tct atc ttc ctc ttg gct ttg      192
195 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
196          50          55          60
198 ctg tcc tgt ctg acc gtt cca gct tcc gct tat gaa gtg cgc aac gtg      240
199 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
200 65          70          75          80
202 tcc ggg atg tac cat gtc acg aac gac tgc tcc aac tca agc att gtg      288
203 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
204          85          90          95
206 tat gag gca gcg gac atg atc atg cac acc ccc ggg tgc gtg ccc tgc      336
207 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys
208          100          105          110
210 gtt cgg gag aac aac tct tcc cgc tgc tgg gta gcg ctc acc ccc acg      384
211 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr

```

## RAW SEQUENCE LISTING

DATE: 08/09/2002

PATENT APPLICATION: US/09/995,860

TIME: 14:42:30

Input Set : A:\SEQUENCE LISTING.txt

Output Set: N:\CRF4\08092002\I995860.raw

```

212          115          120          125
214 ctc gca gct agg aac gcc agc gtc ccc acc acg aca ata cga cgc cac      432
215 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
216          130          135          140
218 gtc gat ttg ctc gtt ggg gcg gct gct ttc tgt tcc gct atg tac gtg      480
219 Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val
220 145          150          155          160
222 ggg gac ctc tgc gga tct gtc ttc ctc gtc tcc cag ctg ttc acc atc      528
223 Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile
224          165          170          175
225 tcg cct cgc cgg cat gag acg gtg cag gac tgc aat tgc tca atc tat      576
226 Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr
227          180          185          190
229 ccc ggc cac ata acg ggt cac cgt atg gct tgg gat atg atg atg aac      624
230 Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn
231          195          200          205
233 tgg tcg cct aca acg gcc ctg gtg gta tcg cag ctg ctc cgg atc cca      672
234 Trp Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
235          210          215          220
237 caa gct gtc gtg gac atg gtg gcg ggg gcc cat tgg gga gtc ctg gcg      720
238 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala
239 225          230          235          240
241 ggt ctc gcc tac tat tcc atg gtg ggg aac tgg gct aag gtt ttg att      768
242 Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile
243          245          250          255
245 gtg atg cta ctc ttt gct ccc taatag      795
246 Val Met Leu Leu Phe Ala Pro
247          260
248 <210> SEQ ID NO: 6
250 <211> LENGTH: 263
251 <212> TYPE: PRT
252 <213> ORGANISM: Hepatitis C virus
255 <400> SEQUENCE: 6
257 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
258 1          5          10          15
260 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
261          20          25          30
263 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
264          35          40          45
266 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
267          50          55          60
269 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
270 65          70          75          80
272 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
273          85          90          95
275 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys
276          100          105          110
278 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr
279          115          120          125

```

## RAW SEQUENCE LISTING

DATE: 08/09/2002

PATENT APPLICATION: US/09/995,860

TIME: 14:42:30

Input Set : A:\SEQUENCE LISTING.txt

Output Set: N:\CRF4\08092002\I995860.raw

```

281 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
282      130      135      140
284 Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val
285 145      150      155      160
287 Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile
288      165      170      175
290 Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr
291      180      185      190
293 Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn
294      195      200      205
296 Trp Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
297      210      215      220
299 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala
300 225      230      235      240
302 Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile
303      245      250      255
305 Val Met Leu Leu Phe Ala Pro
306      260
309 <210> SEQ ID NO: 7
311 <211> LENGTH: 633
312 <212> TYPE: DNA
313 <213> ORGANISM: Hepatitis C virus
315 <220> FEATURE:
316 <221> NAME/KEY: CDS
317 <222> LOCATION: 1..630
319 <220> FEATURE:
320 <221> NAME/KEY: mat_peptide
321 <222> LOCATION: 1..627
323 <400> SEQUENCE: 7
325 atg ttg ggt aag gtc atc gat acc ctt acg tgc ggc ttc gcc gac ctc      48
326 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
327 1      5      10      15
329 atg ggg tac att ccg ctc gtc ggc gcc ccc cta ggg ggt gct gcc aga      96
330 Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
331      20      25      30
333 gcc ctg gcg cat ggc gtc cgg gtt ctg gaa gac ggc gtg aac tat gca      144
334 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
335      35      40      45
337 aca ggg aat ttg cct ggt tgc tct ttc tct atc ttc ctc ttg gct tta      192
338 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
339 50      55      60
341 ctg tcc tgt ctg acc att cca gct tcc gct tat gag gtg cgc aac gtg      240
342 Leu Ser Cys Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
343 65      70      75      80
345 tcc ggg atg tac cat gtc acg aac gac tgc tcc aac tca agc att gtg      288
346 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
347      85      90      95
349 tat gag gca gcg gac atg atc atg cac acc ccc ggg tgc gtg ccc tgc      336
350 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys

```

## VERIFICATION SUMMARY

PATENT APPLICATION: US/09/995,860

DATE: 08/09/2002

TIME: 14:42:31

Input Set : A:\SEQUENCE LSITING.txt

Output Set: N:\CRF4\08092002\I995860.raw

L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date